

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	31	("6859797" "6493703" "6377936" "5923850" "6336139" "20010009017" "20030225763" "20030221163" "6076088" "4270182" "6052693" "6463430" "6601026" "6697793" "6745161"). pn	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2005/11/09 14:30
L2	367	(extract\$3 gather\$3 crawl\$3) NEAR6 (contact NEAR4 (information record\$1))	US-PGPUB; USPAT	OR	OFF	2005/11/09 14:32
L3	132	(extract\$3 gather\$3 crawl\$3) NEAR6 (unstructured NEAR6 (information document\$1))	US-PGPUB; USPAT	OR	OFF	2005/11/09 14:55
L4	22	((extract\$3 gather\$3 crawl\$3) NEAR6 (unstructured NEAR6 (information document\$1))).clm.	US-PGPUB; USPAT	OR	OFF	2005/11/09 14:32
L5	6	2 AND 3	US-PGPUB; USPAT	OR	OFF	2005/11/09 14:36
L6	27	3 AND (contact\$1 NEAR4 (information record\$1))	US-PGPUB; USPAT	OR	OFF	2005/11/09 15:00


[Subscribe \(Full Service\)](#) [Register \(Limited Service,](#)
[Search:](#) [The ACM Digital Library](#) [The Guide](#)
[+extract +unstructured +document +contact +informa](#)
[THE ACM DIGITAL LIBRARY](#)
[Feedback](#) [Report a problem](#) [Satis](#)

Terms used

extract unstructured document contact information

Found

Sort results by [relevance](#)[Save results to a Binder](#)[Try an Advanced Search](#)Display results [expanded form](#)[Search Tips](#)[Try this search in The /](#)☐ [Open results in a new window](#)

Results 1 - 20 of 122

Result page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [next](#)

Relevan

1 [Structured data: Learning to extract information from semi-structured text using](#)[discriminative context free grammar](#)

Paul Viola, Mukund Narasimhan

August 2005 **Proceedings of the 28th annual international ACM SIGIR confer****Research and development in information retrieval SIGIR '05****Publisher:** ACM PressFull text available: [pdf\(178.95 KB\)](#) Additional Information: [full citation](#), [abstract](#), [reference terms](#)

In recent work, conditional Markov chain models (CMM) have been used to extract information from semi-structured text (one example is the Conditional Random Field). Applications range from finding the author and title in research papers to finding the number and street address in a web page. The CMM framework combines a priori encoded as features with a set of labeled training data to learn an efficient extractor. We will show that similar problems can be solved more efficiently.

Keywords: conditional random fields, discriminative grammars, discriminative machine learning, information retrieval, perceptron training, text tagging

2 [Peer-to-peer computing: Guiding queries to information sources with InfoBeacon](#)

Brian F. Cooper

October 2004 **Proceedings of the 5th ACM/IFIP/USENIX international conference on****Middleware****Publisher:** Springer-Verlag New York, Inc.Full text available: [pdf\(315.55 KB\)](#) Additional Information: [full citation](#), [abstract](#), [reference terms](#)

The Internet provides a wealth of useful information in a vast number of dynamic sources, but it is difficult to determine which sources are useful for a given query. Existing techniques either require explicit source cooperation (for example, by exchanging summaries), or build a relatively static source characterization (for example, by analyzing source characteristics).



USPTO

[Subscribe \(Full Service\)](#)
[Register \(Li Free\)](#)

Search: [The ACM Digital Library](#)
[+extract +unstructured +information +pe](#)

THE ACM DIGITAL LIBRARY

[Feedback](#) [Report a problem](#)

Terms used

extract unstructured information people

Sort results by [relevance](#)
☒ [Save results to a Binder](#)

Try an Ad

Display results [expanded form](#)
☒ [Search Tips](#)

Try this search

☐ [Open results in a new window](#)

Results 1 - 20 of 200

Result page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#)

Best 200 shown

1 Ten-year best paper award: Database-inspired search

David Konopnicki, Oded Shmueli

August 2005 **Proceedings of the 31st international conference on VLDB '05**

Publisher: VLDB Endowment

Full text available: [pdf\(309.93 KB\)](#) Additional Information: [full citation](#), [index terms](#)

"W3QL: A Query Language for the WWW", published in 1995, presented distinctive features. Employing existing indexes as access paths, it allows documents using conditions on semi-structured documents and maintain navigational queries. W3QL was capable of automatically filling out forms. Finally, in the SQL tradition, it was a declarative query language, of optimization. Ten years later ...

2 Information extraction

☒ Jim Cowie, Wendy Lehnert

January 1996 **Communications of the ACM**, Volume 39 Issue 1

Publisher: ACM Press

Full text available: [pdf\(645.77 KB\)](#) Additional Information: [full citation](#), [terms, review](#)